Statistical Inference Part 1

Serena Ma

August 18, 2015

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem.Lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials. We will do a thousand simulations.

echo = TRUE #Set the codes visible

Show the sample mean and compare it to the theoretical mean of the distribution.

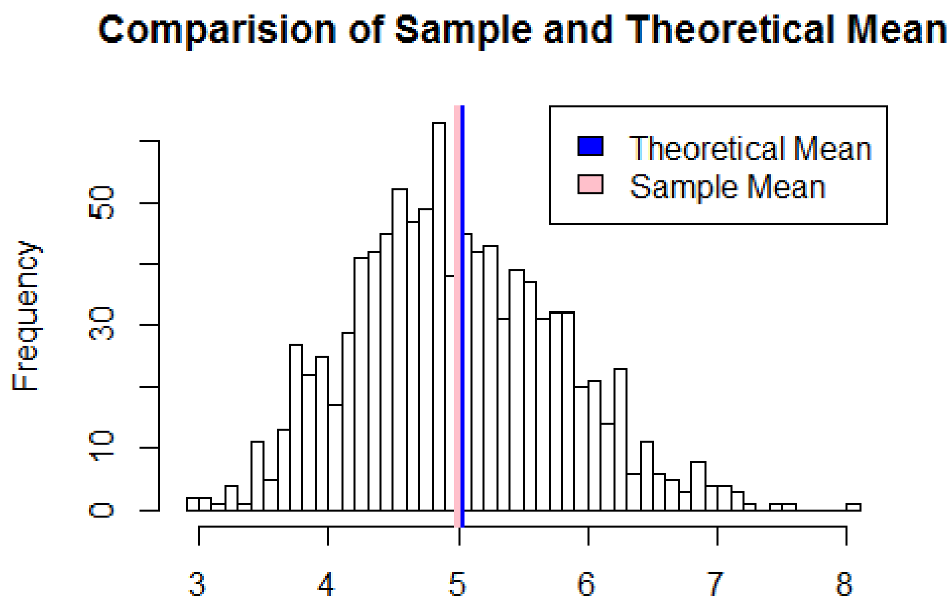
1.Data Simulation:

```
set.seed(1000)
n <- 40
lamda <- 0.2
s <- NULL
for (i in 1:1000) s <- c(s, mean(rexp(n, lamda)))
samplemean <- mean(s)
samplemean

## [1] 4.986963

theoreticalmean <- 1/lamda
theoreticalmean

## [1] 5
```

So the sample mean is 4.986963, and the theoretical mean is 5.

2.Comparison of Sample Mean and Theoretical Mean

```
hist(s, breaks=40, xlab="", main="Comparision of Sample and Theoretical
Mean")
abline(v=theoreticalmean, col="blue", lwd=5)
abline(v=samplemean,col="pink", lwd=3)
legend("topright", c("Theoretical Mean", "Sample Mean"),
fill=c("blue","pink"))
```



From this histogram, we can tell the sample mean and the theoretical mean are very similar. It can also be observed that the disbution of the sample means is quite Gaussian.

Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

1.Data Simulation

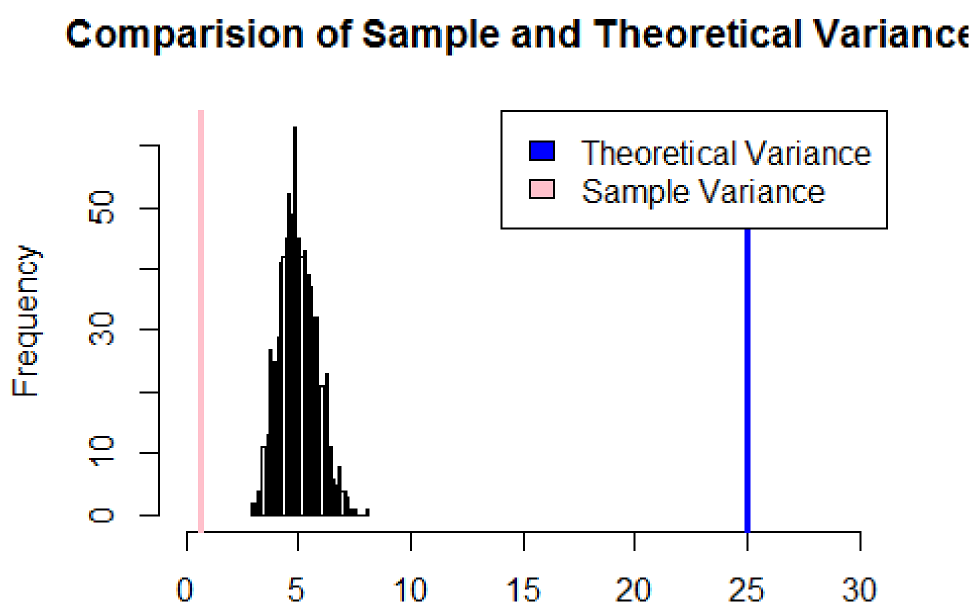
```
samplevariance <- var(s)
samplevariance

## [1] 0.654343

theoreticalvariance <- 1/lamda^2
theoreticalvariance

## [1] 25
```

The sample variance is 0.654343 and the theoretical variance is 25.

In the course we learned that the variance of the sampling distribution of the sample mean is equal to the variance of population (or theoretical) distribution / n. The analysis here holds true. Theoreticalvariance/ sample size n (25/40 = 0.625) is very similar to samplevariance (0.654343).

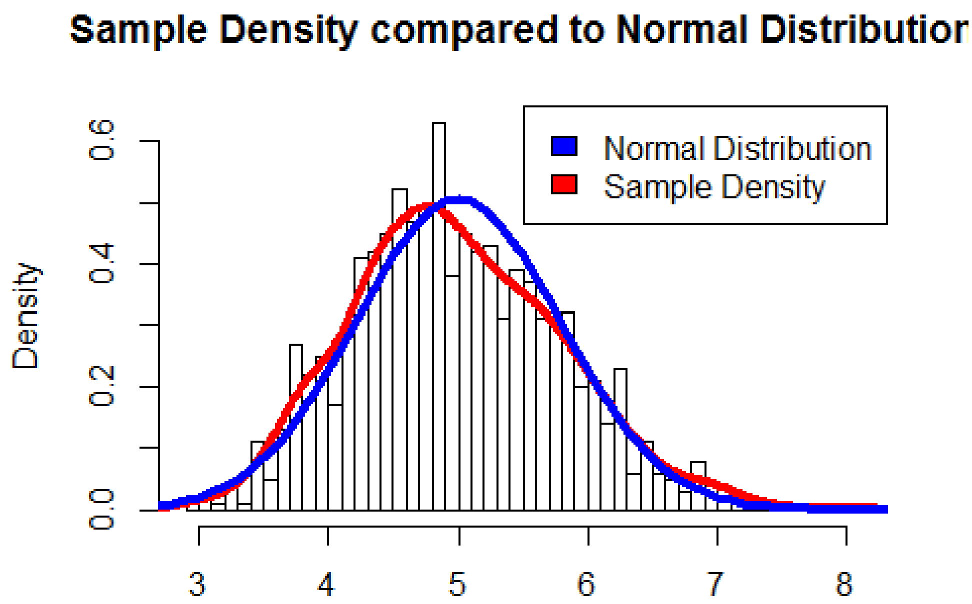
2. Comparision of Sample Variance and Theoretical Variance

```
hist(s, breaks=40, xlim=c(0,30), xlab="", main="Comparision of Sample and
Theoretical Variance")
abline(v=theoreticalvariance, col="blue", lwd=3)
abline(v=samplevariance,col="pink", lwd=3)
legend("topright", c("Theoretical Variance", "Sample Variance"),
fill=c("blue","pink"))
```



Show that the distribution is approximately normal

```
xfit <- seq(1, 10, 0.1)
yfit <- dnorm(xfit, mean = theoreticalmean, sd = sqrt(theoreticalvariance/n))
hist(s, breaks = 40, prob = TRUE,
  xlab = "",
  main = "Sample Density compared to Normal Distribution")
lines(density(s), col = "red", lwd = 4)
lines(xfit, yfit, col="blue", lwd = 4)
legend("topright", c("Normal Distribution", "Sample Density"), fill =
c("blue","red"))
```



From the figure above, we can tell that the sampling distribution of sample means is very close to the normal distribution.